

Reproducibility of Endurance Performance on a Treadmill Using a Preloaded Time Trial

RYAN D. RUSSELL¹, STEPHEN M. REDMANN¹, ERIC RAVUSSIN¹,
GARY R. HUNTER², and D. ENETTE LARSON-MEYER¹

¹Department of Human Studies, Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA; and ²The University of Alabama at Birmingham, Birmingham, AL

ABSTRACT

RUSSELL, R. D., S. M. REDMANN, E. RAVUSSIN, G. R. HUNTER, and D. ENETTE LARSON-MEYER. Reproducibility of Endurance Performance on a Treadmill Using a Preloaded Time Trial. *Med. Sci. Sports Exerc.*, Vol. 36, No. 4, pp. 717–724, 2004.

Purpose: The purpose of this study was to establish a highly reproducible test to measure endurance performance in runners. **Methods:** We evaluated the reproducibility of endurance performance during a 10-km time trial performed on a treadmill after a 90-min preload run at 65% of maximal oxygen uptake ($\dot{V}O_{2\max}$). After screening and a practice test, eight endurance runners (4 men, 4 women, 33.4 ± 10.1 yr, $\dot{V}O_{2\max} = 60.3 \pm 6.3$ mL·kg⁻¹·min⁻¹ in men and 51.8 ± 2.2 mL·kg⁻¹·min⁻¹ in women, mean \pm SD) completed two preloaded time trial tests spaced 3–4 wk apart in men and one menstrual cycle apart in women. A high-carbohydrate diet (15% protein, 10% fat, 75% carbohydrate) was provided the day before both tests. **Results:** Runners completed time trial 1 and time trial 2 in $45:41 \pm 4:45$ and $45:24 \pm 5:03$ min:s, respectively ($43:29 \pm 5:02$ and $43:12 \pm 5:14$ min:s for men and $47:53 \pm 3:47$ and $47:35 \pm 4:23$ min:s for women, trials 1 and 2, respectively). The within-subject coefficient of variation for 10-km time was $1.00\% \pm 0.25\%$ (point estimate \pm estimated standard error) ($0.54\% \pm 0.19\%$ for men and $1.26\% \pm 0.45\%$ for women). **Conclusion:** These results suggest that performance measured as time to complete a 10-km time trial on a treadmill after a 90-min preload is extremely reliable and may be useful for future research assessing the effect of diet, ergogenic substances, or training methods on endurance running performance.

Key Words: RELIABILITY, EXERCISE PERFORMANCE, EXERCISE TEST, FEMALE RUNNERS

The ability to accurately and reliably measure endurance performance is important to the fields of exercise physiology and sports nutrition. Traditionally, endurance performance has been measured in a laboratory while subjects exercise to exhaustion at a set percentage of maximal workload (W_{\max}) or maximal oxygen uptake ($\dot{V}O_{2\max}$) (1,6). However, measures of time to exhaustion (or fatigue) are not directly applicable to real life sports (i.e., running or cycling in a race). Furthermore, such tests have generally not been found to be reproducible measures of endurance performance (2,15,18,21), although one group of researchers used a meta-analysis after converting performance data to mean power to argue to the contrary (14).

Recently, several teams of exercise scientists have developed alternative protocols for measuring endurance performance in the laboratory that should theoretically better replicate the demands of competitive endurance events. These tests have included the measurement of time to complete a given distance (i.e., a race or time trial

(7,10,15,22,24,26) or work completed in a given amount of time (3,15,22,25). Variations to these protocols have included the addition of a submaximal steady-state exercise bout (or preload) before the performance test (7,15) and the inclusion of a series of sprints during the time trial (24). These variations allow for better simulation of the metabolic and/or performance demands of sports such as competitive cycling. Preloaded time trials, in particular, have recently been employed by researchers interested in addressing the effect of substrate altered diets on longer-duration endurance performance (4,5,19). The pretrial preload allows for achievement of a “steady-state,” which may be advantageous to researchers performing isotope studies or measuring gas exchange. If of sufficient duration, preloaded trials also allow for a lowering of endogenous substrate stores before a performance test of higher intensity (7).

Although a number of researchers have found that these alternate performance testing protocols are more reproducible in general than time to exhaustion tests (coefficient of variation (CV) ranging from 0.95 to 4.4, Table 1), most have assessed endurance performance using cycling as the exercise mode (3,10,15,22,24). To our knowledge, only two studies have developed and evaluated a protocol for testing endurance performance during treadmill running (7,25). In these studies, the CV was noted to be slightly higher (CV = 2.7 and 4.4%) than the majority of cycling time trial tests (Table 1) including those conducted by the same investigators (7,24). Similarly, few studies have attempted to determine the reproducibility of performance measures tested with a preload, despite their current popularity. In the few

Address for correspondence: D. Enette Larson-Meyer, Ph.D., RD. Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 80808–4124; E-mail: larsonde@pbrc.edu.

Submitted for publication July 2003.

Accepted for publication November 2003.

0195-9131/04/3604-0717

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 2004 by the American College of Sports Medicine

DOI: 10.1249/01.MSS.0000121954.95892.C8

TABLE 1. Previous studies addressing reproducibility of laboratory studies assessing long duration endurance performance in trained athletes; studies are presented in order of the coefficient of variation (CV).

Authors	Subjects	Mode	Preload	Test	CV	Average Time (min)
Hickey et al. (10)	8 men	Cycle ergometer	No	Time to complete 200 kJ	0.95	12
Hickey et al. (10)	8 men	Cycle ergometer	No	Time to complete 1600 kJ	1.01	105
Palmer et al. (22)	6 men	Kingcycle ergometer	No	Time to complete 40-k time trial	1.0	55
Palmer et al. (22)	6 men	Kingcycle ergometer	No	Time to complete 20-k time trial	1.1	25
Schabert et al. (24)	8 men	Kingcycle ergometer	No	Time to complete 100-k time trial interspersed with 4 1-km and 4 4-km sprints	1.7	156
Smith et al. (26)	8 men	Kingcycle ergometer	No	Time to complete 40-k time trial	1.9–2.1	55
Hickey et al. (10)	8 men	Cycle ergometer	No	Time to complete 14 kJ	2.4	.55
Schabert et al. (25)	8 men	Treadmill	No	Distance completed in 60 min	2.7	60
Bishop et al. (3)	20 women	Cycle ergometer	No	Work completed in 60 min	2.7	60
Jeukendrup (15)	10 men	Cycle ergometer	No	Time to complete set amount of work ($J = 0.75 \times W_{max} \times 3600$)	3.4	60
Jeukendrup (15)	10 men	Cycle ergometer	45 min at 70% $\dot{V}O_{2max}$	Work completed in 15 min	3.5	15
Doyle and Martinez (7)	10 men	Cycle ergometer	90 min at 70% $\dot{V}O_{2max}$	Time to complete a participant specific distance (estimated to take 30 min at preload pace)	3.5	29
Doyle and Martinez (7)	8 men, 2 women	Treadmill	90 min at 70% $\dot{V}O_{2max}$	Time to complete a participant specific distance (estimated to take 30 min at preload pace)	4.4	23
Krebs (18)	10 men	Cycle ergometer	No	Time to exhaustion at 80% $\dot{V}O_{2max}$	20.3	NR
Jeukendrup (15)	10 men	Cycle ergometer	No	Time to fatigue at 75% W_{max}	26.6	61

Kingcycle allows for attachment of cyclists own road bicycle; $\dot{V}O_{2max}$, maximal oxygen uptake; NR, not reported; W_{max} , maximal workload on a cycle ergometer. See Hopkins et al. (14) for a comprehensive review of reliability of power in physical performance.

studies that have been conducted (7,15), including the recent meta-analysis of Hopkins et al. (14), preloaded time trials have been indicated to be inherently less reliable than tests without a preload (7).

The purpose of the present study was to establish a highly reproducible testing protocol to measure endurance performance of longer duration (>2 h total time) in the laboratory in a running population. Specifically, we evaluated the reproducibility of a 10-km time trial on a treadmill that was preceded by a 90-min preload at 65% $\dot{V}O_{2max}$. Similar to the protocol of Doyle and Martinez (7), our protocol was designed to have an initial segment of sufficient duration and intensity to tax endogenous muscle substrate stores (7) and allow for measurement of steady state substrate utilization, followed by a time trial of typical racing distance to assess higher-intensity endurance performance. Such a test, if reproducible, could be used to determine the effect of nutritional and other interventions on performance in endurance runners.

METHODS

The overview of the study protocol and design is shown in Figure 1. The study was approved by Pennington Bio-

medical Research Center (PBRC) Institutional Review Board. All volunteers were briefed about the experimental protocol, and written informed consent was obtained before testing.

Subjects. The subjects were healthy endurance-trained distance runners or triathletes between the ages of 18 and 44 for men, and 18 and 54 for women (in accordance with *ACSM's Guidelines for Exercise Testing and Prescription*) (8). To qualify, participants had to be training regularly (running >32 km·wk⁻¹), had completed at least two training runs of >2 h within the past 3 months, and had a $\dot{V}O_{2max}$ ≥55 mL·kg⁻¹·min⁻¹ for men and ≥50 mL·kg⁻¹·min⁻¹ for women.

Clinic screening. Before admission, volunteers were screened at the PBRC Clinic by a study physician and determined to be in good general health. To qualify, participants also had to have normal fasting insulin, glucose and cholesterol levels, and normal iron status. For descriptive purposes, body composition was also measured using dual-energy x-ray absorptiometry (DEXA, Hologic QDR4500A) (23).

Baseline testing. Before initiation of the experimental protocol, $\dot{V}O_{2max}$ was determined on a motor-driven treadmill (MedTrack ST65, Quinton Industries, Inc., Bothell,

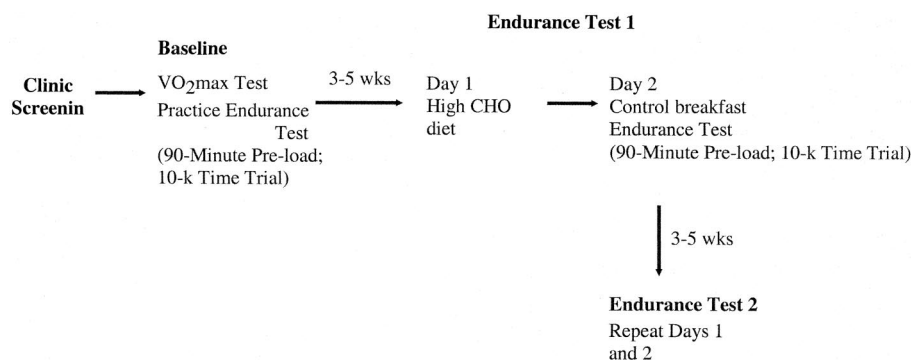


FIGURE 1—Overview of study protocol.

WA) using a previously described protocol designed specifically for the trained running population (20). Briefly, after a 5-min warm-up, “workload” was increased by either speed (starting at the subject’s typical warm-up speed and increasing by 0.5 mph (13.4 m·min⁻¹)) or grade (starting at 0° and increasing by 2.5%) every minute until exhaustion. The increase in workload was selected by the participant 15–20 s before completion of the stage. Oxygen consumption ($\dot{V}O_2$) and carbon dioxide production ($\dot{V}CO_2$) were measured using a metabolic cart (V-Max 29 Series, Sensor-Medics, Yorba Linda, CA) that was calibrated before each test. Heart rate was also monitored continuously using a portable heart rate monitor (Polar S-610, Polar Beat, Port Washington, NY). The highest $\dot{V}O_2$, respiratory exchange ratio (RER), and heart rate achieved over a 20-s period within the last 2 min of exercise were recorded as the maximum values. For the test to be considered an acceptable measurement of $\dot{V}O_{2max}$, two of the following criteria had to be met: 1) a leveling or plateau of $\dot{V}O_2$ (defined as an increase of $\dot{V}O_2$ of < 2 mL·kg⁻¹·min⁻¹ with increased workload), 2) RER > 1.10, and 3) maximum heart rate within 10 beats of age-predicted maximum (max heart rate = 208 – (0.7)(age) as the predicting equation) (27).

After the $\dot{V}O_{2max}$ test, subjects who met the $\dot{V}O_{2max}$ criteria for admission into the study rested for 30 min and then performed a practice 90-min preload and 10-km time trial using the procedures described below for endurance test 1 and endurance test 2. The purpose of the practice test was to familiarize the subjects with the treadmill and the testing equipment, and to reduce a potential learning curve from the first to the second trials (24). Food and fluid intake, however, were not controlled before the practice test and the results were not treated to statistical analysis.

Experimental protocol. As outlined in Figure 1, participants completed two separate endurance tests that were spaced approximately 3–4 wk apart in men and one menstrual cycle (3–5 wk) in women. The endurance test consisted of a 90-min preload run at 65% $\dot{V}O_{2max}$ followed by a 10-km time trial. Diet was controlled by providing volunteers with a high-carbohydrate, low-fat diet (15% protein, 10% fat, 75% carbohydrate) and a controlled breakfast (15% protein, 15% fat, 70% carbohydrate) before both endurance tests. Menstrual cycle phase (follicular or luteal) was estimated by the start of the subject’s last menstrual period and documented by serum estradiol and progesterone levels from a blood sample drawn the morning of the test. Participants were asked to keep their diet and training consistent between trials.

Endurance test. The endurance performance test was performed in the morning, 3 h after the subject consumed the controlled 500-kcal breakfast. The 90-min preload run and 10-km time trial were performed on the same laboratory treadmill (LifeFitness TR 9500, Franklin Park, IL). Before the 90-min preload, a 5-min warm-up was performed at a self-regulated pace and was kept consistent between trials. For most participants, an incline of 1% was assigned during both the preload and time trial runs to assimilate the oxygen cost of outside running (16). However, due to limitations of

our treadmill (maximal velocity = 266.67 m·min⁻¹), faster volunteers (those who had completed recent 10-km runs in less than 40 min) were given a higher incline of between 2 and 3% during all 10-km time trials. Heart rate was monitored continuously during the preload and time trial runs using a portable heart rate monitor (Polar S-610, Polar Beat, Port Washington, NY). $\dot{V}O_2$ and $\dot{V}CO_2$ were measured for 5 min at the beginning (between 5 and 15 min after the start) and end (10 min before completion) of the 90-min preload to verify the oxygen cost (as explained below), and to measure the respiratory quotient (RQ). Gas exchange was not measured during the 10-km time trial so as not to interfere with performance.

During the 90-min preload, participants ran at a steady pace, which elicited a $\dot{V}O_2$ of 65% of $\dot{V}O_{2max}$. This pace was initially determined during the baseline practice trial and adjusted (if necessary) at the beginning of test 1. During the time trial run, participants were instructed to complete the 10-km run in the least amount of time as possible (i.e., to race the 10-km) and were allowed to adjust the treadmill speed as much as desired. Subjects were given feedback with respect to distance and time elapsed only at each mile mark (mile splits), and hence only knew their time at each mile they had passed at that point. Mile splits were documented and used to determine pacing patterns. Subjects were informed as to what their final 10-km time was at the completion of each 10-km. Water consumption was encouraged and allowed *ad libitum* (in weighed water bottles) during the first endurance test (as well as the practice tests) and was then replicated during the second session. Body mass was taken before and after each run, both before and after the participants voided. Subjects were not allowed to consume anything other than water during each of the tests.

Volunteers were allowed to watch a movie, listen to music, or both during each testing session. However, the same mode of entertainment employed for the first test was required for the second test (i.e., same movie or music CD), and the same researcher that monitored the first test also monitored the second test. Although highly motivated endurance-trained athletes were recruited, additional prizes of interest to runners were given during the 10-km portion of performance tests 1 and 2 as external motivators. Specifically, volunteers were promised a workout towel for coming within 90 s of their 10-km personal record (PR) (set at any competitive 10-km event over the last 2 yr), and a runner’s PR package (\$50 gift certificate to a local running store, and certificate for $\dot{V}O_{2max}$ testing and body composition analysis at PBRC at a future date) for meeting or exceeding their PR. PR times were verified from race records.

Diet and pretrial breakfast. The high-carbohydrate, low-fat diet, and pretrial control breakfast provided the day before the endurance tests was prepared by the PBRC Metabolic Kitchen, and consisted of commercially available foods and beverages. Eucaloric requirements were estimated from the Harris Benedict prediction equation for basal energy expenditure (BEE) (9) multiplied by an activity factor and rounded to the nearest 100 kcal. An activity factor of 2.0 was used for male runners and an activity factor of 1.8 was used for

female runners. Volunteers ate breakfast and dinner at the center, and were given lunch and snacks to carry out. The pretest control breakfast was the same for all volunteers, and provided 500 kcal (15% protein, 15% fat, 70% carbohydrate; average glycemic index ~ 62). If the subjects were habitual consumers of caffeine in the form of coffee, tea, or cola, a similar amount of caffeine as measured instant coffee powder, instant tea powder, or diet cola was provided with their test breakfast. Volunteers were asked to eat all food given and to return anything not consumed to the research dietitian so it could be weighed and recorded. The pretest diet and control breakfast for testing session 2 was adjusted if any food was returned during testing session 1. Before both endurance tests, volunteers were also asked to hydrate in the 24-h period before testing by striving to consume about 4536 g of water.

Statistical approach. The primary response variable in this reproducibility study is time to complete the 10-km time trial in minutes (T_e) after the timed preload exercise test. The CV that is reported throughout this paper is based on the ratio of the square root of the estimated within-subject component of the total variance of T_e to the estimate of the population mean. This CV is a measure of repeatability or reproducibility (11) and is often identified as the within-subject CV. A second measure of the repeatability of the measurements is the intraclass correlation (ICC), which is the ratio of the between-subjects component of the variance to the total variance of T_e . Estimates of the ICC that are close to unity (1.0) reflect the fact that measurement error is small relative to the total variance.

Random effects models were used to first estimate both the between-subject and the within-subject components of the total variance, along with the mean. These parameter estimates were then employed to estimate the within-subject coefficient of variation and the ICC for each group (males and females). The within-subject CV is given as $CV = 100(s_w)/m_T$, where s_w is the square root of the estimate of the within-subject variance component for a particular sex, and m_T is the estimate of the population mean for 10-km time for that group. To obtain an estimate of the standard error of this statistic in terms of the parameter estimates obtained from the model and their asymptotic variance estimates, a first-order Taylor series was employed to derive an approximating expression. The parameter estimates from the model and their asymptotic variance estimates were substituted into this expression to obtain the estimated standard error of the CV. The ICC is estimated as: $s_b^2/(s_b^2 + s_w^2)$, where s_b^2 is the estimated between-subjects component of the variance.

To estimate the ICC and the within-subject CV for both genders overall, the data were aggregated and a single random effects model was used to estimate the population mean and the two variance components. These parameter estimates were used in a similar fashion to estimate the overall ICC and the within-subject CV.

RESULTS

Ten participants were initially enrolled in the study; however, two male participants were unable to complete the

TABLE 2. Subject characteristics.

Characteristics	Men (N = 4)	Women (N = 4)
Age (yr)	34.6 \pm 11.2	36.1 \pm 11.1
Height (cm)	178.6 \pm 7.6	164.8 \pm 2.6
Weight (kg)	69.8 \pm 10.7	56.5 \pm 7.6
Body fat (% by DEXA)	12.4 \pm 2.0	19.6 \pm 2.6
$\dot{V}O_{2max}$ (mL·kg ⁻¹ ·min ⁻¹)	60.3 \pm 6.3	51.8 \pm 2.2
$\dot{V}O_{2max}$ (mL·kg FFM ⁻¹ ·min ⁻¹)	68.7 \pm 2.0	64.4 \pm 2.6
Maximum heart rate (bpm)	188 \pm 4	184 \pm 8

Mean \pm standard deviation; DEXA, dual energy x-ray absorptiometry; $\dot{V}O_{2max}$, maximal oxygen uptake; FFM, fat-free mass; bpm, beats per minute.

study. One participant was not able to complete the first 10-km time trial due to severe fatigue. The second sustained an overuse injury during the study. Though the second volunteer finished the study, he complained of severe knee pain during the endurance test and ended up completing the 10-km time trial in a time that was almost 20 min slower than his practice time. Both volunteers were therefore not included in the data analysis.

The characteristics for the eight athletes who completed the study (four men and four women) are shown in Table 2. All participants were well-trained recreational runners who reported running an average of 62.0 \pm 1.9 km·wk⁻¹ (59.5 \pm 13 km for men, 64.4 \pm 15.7 km for women). Most competed regularly in local road races including 10-km, marathon, and ultra marathon distances. Participants reported consuming all of the food/beverages provided in the pretest diet which averaged 2736 \pm 452 kcal·d⁻¹ for the group (3050 \pm 387 kcal·d⁻¹ for the men and 2400 \pm 216 kcal·d⁻¹ for women). The endurance test in sessions 1 and 2 were performed in the same phase of the menstrual cycle in all women; however, two were studied in the follicular phase (progesterone < 0.2 ng·mL⁻¹) and two were studied in the luteal phase (progesterone range = 5.7 to 9.8 ng·mL⁻¹).

During the 90-min preload, subjects ran at an average pace of 3.12 \pm 0.44 m·s⁻¹, which elicited an average oxygen cost of 64.0 \pm 4.0 and 63.8 \pm 3.4 percentage of $\dot{V}O_{2max}$ during test 1 and test 2, respectively. When stratified by gender, men ran at an average pace of 3.43 \pm 0.51 m·s⁻¹, which elicited an oxygen cost of 64.2 \pm 4.2 and 63.6 \pm 3.4% $\dot{V}O_{2max}$ during tests 1 and 2, respectively, and women ran at an average pace of 2.87 \pm 0.33 m·s⁻¹, which elicited an oxygen cost of 63.8 \pm 4.62 and 63.2 \pm 4.2% $\dot{V}O_{2max}$ during tests 1 and 2, respectively. Table 3 summarizes the RQ, heart rate, water intake, and body mass changes associated with the 90-min preload runs. These data illustrate that the preload conditions before time trials 1 and 2 were similar.

As shown in Figure 2, the group of runners completed the 10-km time trials in 45:41 \pm 4:45 and 45:24 \pm 5:03 min:s. The average 10-km times by gender were 43:29 \pm 5:02 and 43:12 \pm 5:14 min:s for men and 47:53 \pm 3:47 and 47:35 \pm 4:23 min:s for women on trials 1 and 2, respectively. Paired *t*-tests indicate that there was insufficient evidence to conclude that a difference exists in mean elapsed time (10-km time) between trials 1 and 2, either for the overall population ($P = 0.23$) or for the gender groups ($P = 0.07$ and 0.57 , for men and women, respectively). For the group, heart rate

TABLE 3. Respiratory quotient (RQ), heart rate (HR), water consumption, and change in body mass during the 90-min preload run before time trials 1 and 2.

	Trial 1		Trial 2	
	Men	Women	Men	Women
RQ				
Beginning	0.95 ± 0.03	0.92 ± 0.03	0.94 ± .04	0.93 ± 0.04
End	0.92 ± 0.03	0.93 ± 0.04	0.91 ± .02	0.94 ± 0.03
Heart rate				
Beginning (bpm)	144 ± 11	145 ± 12	140 ± 12	142 ± 11
End (bpm)	148 ± 12	146 ± 10	149 ± 11	146 ± 12
Average (bpm)	147 ± 11	145 ± 13	143 ± 9	147 ± 14
Average (% max)	78 ± 7	79 ± 5	76 ± 6	80 ± 5
Water consumed (g)	910 ± 636	432 ± 236	822 ± 604	481 ± 207
Body mass change (kg) ^a	-0.9 ± 0.5	-0.5 ± 0.5	-1.0 ± 0.2	-0.4 ± 0.3

Mean ± standard deviation for men and women: bpm, beats per minute.

^aBody mass change is calculated as the difference between prerun body mass and postrun body mass before voiding.

during the 10-km run averaged 173 ± 6 and 175 ± 6 bpm (94 ± 3 and $95 \pm 4\%$ of maximum heart rate) on trials 1 and 2, respectively. Average 10-km heart rate by gender was 175 ± 5 and 175 ± 5 bpm for men (93 ± 4 and 94 ± 2 percentage of maximum heart rate) and 170 ± 7 and 175 ± 8 bpm for women (94 ± 2 and 95 ± 6 percentage of maximum heart rate) on trials 1 and 2, respectively.

The within-subject CV for time to complete the 10-km time trial was $1.00\% \pm 0.25\%$ (estimate ± standard error) for the group of men and women, and the ICC was 0.991 (Fig. 3). When calculated separately by gender, the CV was $0.54\% \pm 0.19\%$ for men and $1.26\% \pm 0.45\%$ for women. The ICC for men was 0.998, and for women it was 0.985.

The pacing patterns of the male and female participants during the 10-km run are shown in Figure 4a and b. The solid lines are the subjects' mean time for the two trials, and the size of the square placed at each mile point is proportional to the difference for that subject at that mile marker across trials (i.e., the size of the symbols represent within subject variability). These figures illustrate near-linear pacing patterns in these trained runners and do not indicate pace slowing or pace acceleration across the miles. (Note, times at each mile rather than time at each kilometer were documented because this typically occurs in road races in the United States). The figures also indicate that the within subject difference across all mile points are smaller for male compared to female runners in our sample.

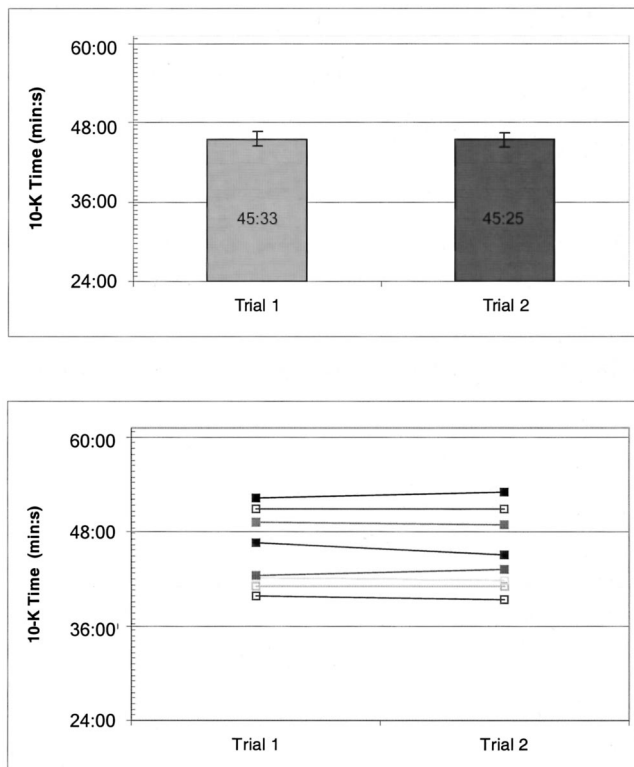


FIGURE 2—Time to complete the 10-km time trial 1 and time trial 2 for the group of runners (top panel, mean ± SEM) and for individual male (open squares) and female (closed squares) runners (bottom panel). Trials 1 and 2 were completed after a 90-min preload run at $65\% \dot{V}O_{2max}$.

DISCUSSION

The first important finding of the present study was that a laboratory simulated 10-km performance run on a treadmill following a 90-min preload run (at $65\% \dot{V}O_{2max}$) designed to compromise muscle substrate stores was highly reproducible, as indicated by the small within subject CV. The CV for time to complete a 10-km time was 1% when the

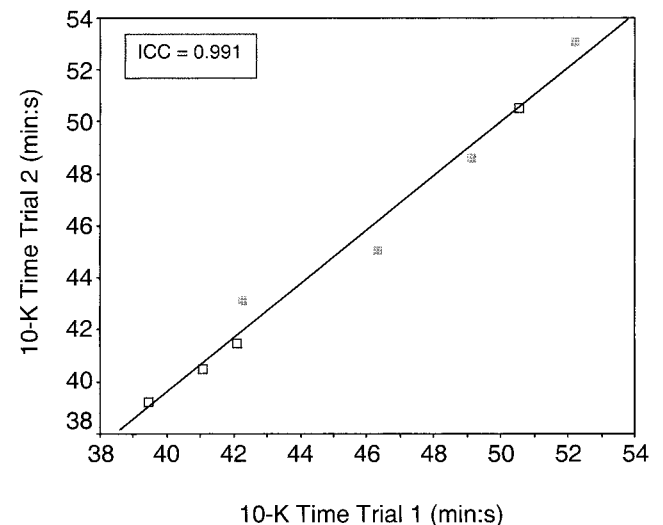


FIGURE 3—Correlation between 10-km time on time trials 1 and 2. Data for men are shown in open squares. Data for women are shown in closed squares. The intraclass correlation estimate (ICC) was 0.991.

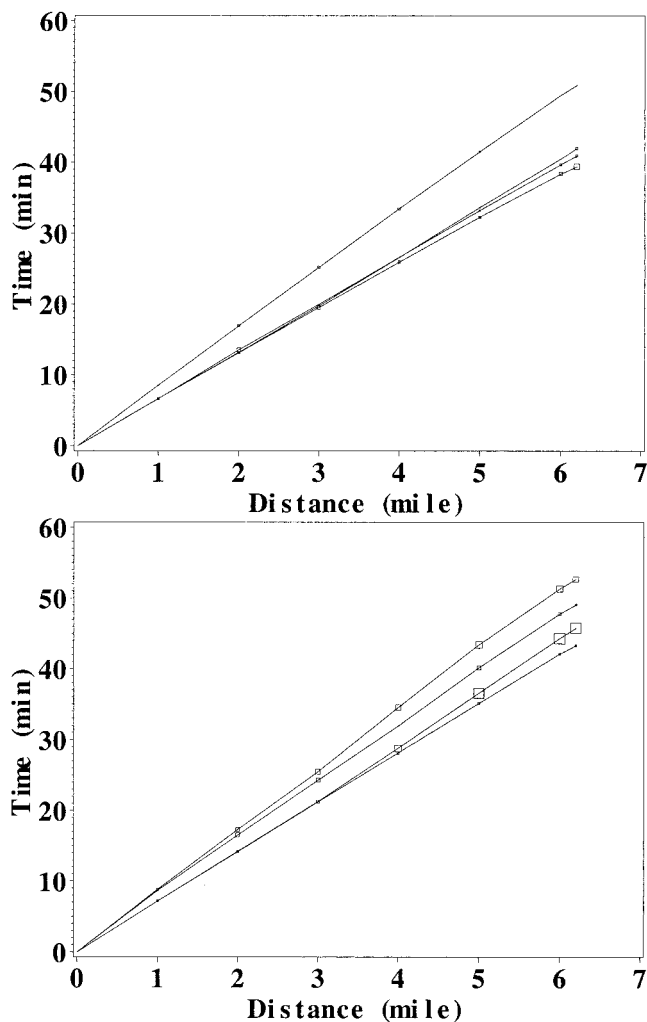


FIGURE 4—Time at mile points for four male (upper panel) and female (lower panel) runners. Symbol height is time difference between trials.

data for both genders (four men and four women) were aggregated and was slightly better in the male ($CV = 0.54\%$) than the female ($CV = 1.26\%$) runners. Other interesting findings were that the pacing patterns during the 10-km runs were linear and appeared to be slightly more variable across all mile points in female compared with male runners. Overall, the results of this study indicate that a running time trial test can be used as a reliable method for assessing endurance performance in well-trained male and female runners.

While the CV for performance time in our preloaded time trial was quite impressive, it is in line with published studies assessing the reliability of performance time in similarly trained cyclists performing time trial events of comparable time or distances (10,15,22) of up to 100 km (24). For instance, Hickey et al. (10) determined that the CV of time to complete four successive performance rides of 1600 kJ (which took an average of 105 min) on a computerized cycle ergometer was 1.01% in 10 competitive male cyclists. Palmer et al. (22) observed that the reproducibility of performance time during three simulated 40-km time trials

(which took an average of ~ 56 min) in six competitive cyclists who performed on their own bikes attached to a Kingcycle bike stand was 1.0%. Researchers in both studies, however, did not standardize diet or fluid intake before the time trials as was done in the current study.

With respect to running studies, however, the reproducibility of time to complete a 10-km time trial (after a 90-min preload) in the current study (which took an average of ~ 45 min) was better than that reported by the two previous studies assessing performance during treadmill running (7,25). In these studies, Schabort et al. (25) found that the CV for distance completed during three 60-min self-paced runs on a treadmill in male runners was 2.7%, whereas Doyle and Martinez (7) found that the CV for performance time during a 90-min preloaded time trial of participant-specific distance in male and female runners was 4.4% when the first of the four trials was considered a practice trial ($CV = 10.1\%$ for all four trials). Reasons for better reproducibility in our trial are not clear but may be related to the performance testing protocol employed, which in the present study was a time trial of a distance common to road-racing (i.e., 10-km). Although Doyle and Martinez (7) also conducted a similar preloaded time trial, the distance covered varied by participant and averaged 5.05 km. In support of this possibility, Jeukendrup et al. (15) speculated from their studies in well-trained cyclists that motivation to finish faster in a race-like setting may favor more reproducible results in a time trial protocol (15), which could be expected to be more reliable when distances typical of racing events are used. An additional possibility is that the current study more tightly regulated the diet before and during the test-retest trials. Whereas we fed participants an isocaloric high-carbohydrate diet for 24-h before testing, provided a standardized breakfast, and standardized fluid intake (session 2 to session 1), both Schabort et al. (25) and Doyle and Martinez (7) instructed participants to maintain the same diet regimen for 48 h before each run based on a set of guidelines and allowed *ad libitum* water intake during all trials. Doyle and Martinez (7) also had participants performed their trials in the fasting state, which may have contributed to the greater variable noted in their study. Although subject characteristics can also potentially contribute to differences in performance reproducibility (13), differences between the present and previous studies are unlikely to explain the better reproducibility of our study. For example, the runners included in the studies of Schabort et al. (25) and Doyle and Martinez (7) were on average slightly younger (7,25) and faster (25) than the male runners in our study. Based on previous finding, however, older runners and faster runners would be expected to be less variable (more reproducible) in their performance than younger runners and slower runners (13).

Findings from the present study are important because they suggest that treadmill running performance during a preloaded time trial test was not inherently less reliable than cycling performance (25) or performance measured without a preload (14) as has been previously suggested. In fact, we found that the reproducibility of 10-km time on a treadmill

after a preload run was less variable than performance time during actual races and fell within the acceptable range suggested by Hopkins and Hewson (13) for measuring performance in the laboratory. In their studies assessing the variability of competitive performance in distance runners, Hopkins and Hewson (13) found that the CV for official race times for males in the faster half was 1.7% for the 8-km cross country race and between 2.9 and 4.7% for the half marathon. They also speculated that tests of endurance performance suitable for assessing the smallest worthwhile changes in running performance for top runners (assumed to be 0.5% for a cross-country or road race and 1.0% for a marathon or half marathon (12)) require a CV of $\leq 1.5\%$ for tests simulating a typical cross-country distance and $\leq 2.5\%$ for those simulating a half or full marathon (13). Thus, our preloaded 10-km time trial should be reliable enough to detect any meaningful differences in running performance as a result of a research intervention without the need for a sample size that is beyond the resources or pool of subjects available to the researcher (12).

Also of interest in the present study was the finding that endurance performance was quite reproducible in male and female runners alike. This is important because the majority of studies have assessed performance reproducibility in male athletes with only a few assessing this relationship in exclusively female (3) or mixed male and female participants (7). The most comprehensive studies in the performance arena were done by Hopkins and Hewson (13) in the field where they addressed the typical variation in competitive running performance during races of a variety of distances. They found that times of male runners were less reproducible in general than those of the female runners. In our study, we found both that the point estimate for the within-subject CV for male subjects was smaller, perhaps negligibly so, than that for female subjects (0.54% for male runners and 1.26%, for female runners) and that the within subject difference across all mile points (pacing pattern) was also smaller (and therefore less variable) for the male compared to the female runners in our sample. Although this might be suggestive of a contradiction of the findings of Hopkins and Hewson, it must be borne in mind that there is no practical implication of the very small apparent difference between the two estimates; both point estimates are small in magnitude and therefore are indicative of highly repeatable measures. There, however, are unresolved statistical questions regarding the distribution of the CV estimators, particularly for small sample sizes, that preclude our testing for gender differences. Further studies will be needed to help resolve this issue.

In the laboratory setting, we like others (15,19) had the athletes undergo a practice trial in order to familiarize them with treadmill running and minimize any possible learning effect (7,14,24), which may be particularly important during treadmill running (7). Despite employment of this practice trial, however, we found that six of the eight volunteers (three of each gender) ran faster during the second time trial. Although this difference was not statistically significant in the current study, it renders attention to the importance of

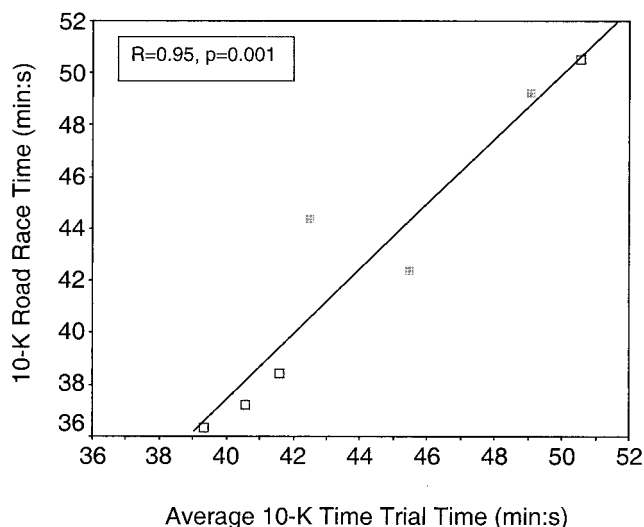


FIGURE 5—Relationship between average time to complete the 10-km time trial in the laboratory (after the 90-min preload run at 65% $\dot{V}O_{2max}$) and time in a recent 10-km road race ($N = 7$). Data for men are shown in open squares. Data for women are shown in closed squares.

treatment randomization when a possible performance effect is of interest. The results of Hickey et al. (10) nicely illustrate a similar phenomenon or the presence of what could be called a “last trial” effect during longer time trials where athletes simply perform better on their final trial. In their studies, competitive cyclists asked to perform four successive trials at three workloads (14, 200, and 1600 kJ) designed to mimic a short (~30 s), medium (~12 min), and long (~105 min) time trials were found to perform significantly faster during the fourth and final trial compared with the previous three trials. A last trial effect, however, has not been detected by all investigators (15,22).

A final point of interest to endurance performance testing is whether a test of endurance performance simulates the demands of the endurance event and is a valid measurement of endurance performance. Although the current study was not designed to test validity, we compared the time trial results collected in the lab with race results from a large local 10-km competition that occurred shortly after the participants completed the laboratory trials (data not reported in results). Quite interestingly, we found a strong correlation ($r = 0.95$) between average performance time during the 10-km time trial on the treadmill and 10-km road time in seven of the eight runners who participated in the race (Fig. 5). The participants’ average 10-km time in the laboratory was 1.6 min slower than their recent 10-km road race times, which is not surprising given that the participants ran for 90 min before completing the time trial and were thus racing with compromised endogenous stores. Two of the faster participants, in addition, were given a slight grade during the treadmill run. In the study of competitive cyclists by Palmer and colleagues (22), 40-km time trial time in the laboratory on a Kingcycle system was also found to correlate ($r = 0.98$) with 40-km time trial time during a road competition. Road race time was an average 8% slower,

which is likely due to slowing by wind resistance and other environmental factors. In the current study, however, it is interesting that our pacing patterns did not indicate pace slowing across the 10-km run as may have been expected in response to the predicted glycogen depletion (17) induced by the 90-min preload run (average distance covered = 16.9 km). One possible explanation for this finding is that pace was slowed below optimal from the beginning of the 10-km run (an estimated $9.6 \text{ s}\cdot\text{km}^{-1}$ slower based on 10-km road race time), indicating possibly that these well-trained runners were able to sustain a reproducible and steady suboptimal pace by shifting to fat oxidation as the primary source of fuel. Clearly, future studies using this preloaded protocol

combined with measures of whole-body and muscle substrate utilization should provide some interesting insights on this issue.

In conclusion, we found that endurance performance measured during a preloaded time trial on a treadmill is extremely reproducible when measured in endurance-trained male and female runners. This type of testing may be important for future research assessing the effect of diet, ergogenic substances, or training methods on endurance running performance.

This work was partially supported by U01 AG20478 and K01 DK062018-01.

REFERENCES

- BERGSTROM, J., L. HERMANSEN, E. HULTMAN, and B. SALTIN. Diet, muscle glycogen and physical performance. *Acta Physiol. Scand.* 71:140–150, 1967.
- BILLAT, V., J. C. RENOUX, J. PINOTEAU, B. PETIT, and J. P. KORALSZTEIN. Reproducibility of running time to exhaustion at $\text{VO}_{2\text{max}}$ in subelite runners. *Med. Sci. Sports Exerc.* 26:254–257, 1994.
- BISHOP, D. Reliability of a 1-h endurance performance test in trained female cyclists. *Med. Sci. Sports Exerc.* 29:554–559, 1997.
- BURKE, L. M., D. J. ANGUS, G. R. COX, et al. Effect of fat adaptation and carbohydrate restoration on metabolism and performance during prolonged cycling. *J. Appl. Physiol.* 89:2413–2421, 2000.
- BURKE, L. M., J. A. HAWLEY, D. J. ANGUS, et al. Adaptations to short-term high-fat diet persist during exercise despite high carbohydrate availability. *Med. Sci. Sports Exerc.* 34:83–91, 2002.
- CHRISTENSEN, E. H., and O. HANSEN. II. Untersuchungen über die Verbrennungs-Vorgänge bei Langdauernder, schwerer Muskelarbeit. III. Arbeits Fähigkeit und Ernährung. IV. Hypoglykämie, Arbeitsfähigkeit und ermüdung. [Research into the oxidative processes during heavy endurance workouts (muscle involvement). III. Work capacity and nutrition. IV. Hypoglycemia, work capacity and exhaustion.] *Skand. Arch. Physiol.* 81:153–181, 1939.
- DOYLE, J. A., and A. L. MARTINEZ. Reliability of a protocol for testing endurance performance in runners and cyclists. *Res. Q. Exerc. Sport* 69:304–307, 1998.
- FRANKLIN, B. A. *ACSM's Guidelines for Exercise Testing and Prescription*, 6th Ed., M. H. Whatley (Ed.). Baltimore: Lippincott Williams & Wilkins, 2000, p. 368.
- HARRIS, J., and F. BENEDICT. *A Biometric Study of Basal Metabolism in Man*. Washington, DC: Carnegie Institute of Washington, 1919, pp. 227.
- HICKEY, M. S., D. L. COSTILL, G. K. MCCONELL, J. J. WIDRICK, and H. TANAKA. Day to day variation in time trial cycling performance. *Int. J. Sports Med.* 13:467–470, 1992.
- HOPKINS, W. G. Measures of reliability in sports medicine and science. *Sports Med.* 30:1–15, 2000.
- HOPKINS, W. G., J. A. HAWLEY, and L. M. BURKE. Design and analysis of research on sport performance enhancement. *Med. Sci. Sports Exerc.* 31:472–485, 1999.
- HOPKINS, W. G., and D. J. HEWSON. Variability of competitive performance of distance runners. *Med. Sci. Sports Exerc.* 33:1588–1592, 2001.
- HOPKINS, W. G., E. J. SCHABORT, and J. A. HAWLEY. Reliability of power in physical performance tests. *Sports Med.* 31:211–234, 2001.
- JEUKENDRUP, A., W. H. SARIS, F. BROUN, and A. D. KESTER. A new validated endurance performance test. *Med. Sci. Sports Exerc.* 28:266–270, 1996.
- JONES, A. M., and J. H. DOUST. A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *J. Sports Sci.* 14:321–327, 1996.
- KARLSSON, J., and B. SALTIN. Diet, muscle glycogen, and endurance performance. *J. Appl. Physiol.* 31:203–206, 1971.
- KREBS, P. S., and S. K. POWERS. Reliability of laboratory endurance tests. *Med. Sci. Sports Exerc.* 21:510, 1989.
- LAMBERT, E. V., J. H. GOEDECKE, C. ZYLE, et al. High-fat diet versus habitual diet prior to carbohydrate loading: effects of exercise metabolism and cycling performance. *Int. J. Sport Nutr. Exerc. Metab.* 11:209–225, 2001.
- LARSON-MEYER, D. E., B. R. NEWCOMER, and G. R. HUNTER. Influence of endurance running and recovery diet on intramyocellular lipid content in women: a ^1H NMR study. *Am. J. Physiol. Endocrinol. Metab.* 282:E95–E106, 2002.
- MCLELLAN, T. M., S. S. CHEUNG, and I. JACOBS. Variability of time to exhaustion during submaximal exercise. *Can. J. Appl. Physiol.* 20:39–51, 1995.
- PALMER, G. S., S. C. DENNIS, T. D. NOAKES, and J. A. HAWLEY. Assessment of the reproducibility of performance testing on an air-braked cycle ergometer. *Int. J. Sports Med.* 17:293–298, 1996.
- PRITCHARD, J. E., C. A. NOWSON, B. J. STRAUSS, J. S. CARLSON, B. KAYMAKCI, and J. D. WARK. Evaluation of dual energy X-ray absorptiometry as a method of measurement of body fat. *Eur. J. Clin. Nutr.* 47:216–228, 1993.
- SCHABORT, E. J., J. A. HAWLEY, W. G. HOPKINS, I. MUJKA, and T. D. NOAKES. A new reliable laboratory test of endurance performance for road cyclists. *Med. Sci. Sports Exerc.* 30:1744–1750, 1998.
- SCHABORT, E. J., W. G. HOPKINS, and J. A. HAWLEY. Reproducibility of self-paced treadmill performance of trained endurance runners. *Int. J. Sports Med.* 19:48–51, 1998.
- SMITH, M. F., R. C. DAVISON, J. BALMER, and S. R. BIRD. Reliability of mean power recorded during indoor and outdoor self-paced 40 km cycling time-trials. *Int. J. Sports Med.* 22:270–274, 2001.
- TANAKA, H., K. D. MONAHAN, and D. R. SEALS. Age-predicted maximal heart rate revisited. *J. Am. Coll. Cardiol.* 37:153–156, 2001.